RESEARCH ON CRIMES IN SMALL AREAS: TRAPS AND TRICKS*

by

Dennis W. Roncek, Ph.D.
Department of Criminal Justice
University of Nebraska at Omaha

Paper presented to The Seventh Annual International Crime Mapping Research Conference in Boston, Massachusetts in April 2004 sponsored by the National Institute of Justice of the United States Department of Justice. The views and opinions presented in this paper are solely those of the author and do not necessarily represent the positions of any governmental agency at the local, state, or national level. As always, any remaining faults are the sole responsibility of the author.

RESEARCH ON CRIMES IN SMALL AREAS: TRAPS AND TRICKS

Dennis W. Roncek, Ph.D
Department of Criminal Justice
University of Nebraska at Omaha

Abstract

As concern with the locations of crime, criminals, and victims has increased, the importance of accurately identifying these locations has become much more serious. In large part, this concern with precision has increased due to the importance of linking advanced research methods with policy efforts. This workshop identified several advantages and pitfalls of working with small area data. It discussed the availability of data from the Census, strategies for overcoming the gaps in data from the Census, the problems of working with agency data for small areas, along with methodological and statistical strategies to overcome these problems.

INTRODUCTION:

Three Fundamental Tricks to Avoid Traps

Before diving into the depth of issues about working with small areas or even how these will be defined for this discussion, there are three tricks that will be invaluable in avoiding the traps of working with small area data. These are given below.

The first major trick to paraphrase a former resident of Boston is "Ask not what you should do because the software will do it, but ask which software will do what should be done." I have read many manuscripts in which a particular piece of research was done simply because the software (and those who are familiar with my working style know which I mean) can do a particular task not because there was anything that could be really gained from it. Be multiplatform, have no loyalty to any software for mapping or for statistical analysis. It may cost a bit more, but it will be possible to do so much more.

The second trick is to keep multiple copies of all maps, but this is especially important for the street maps for which a nice presentation copy will be needed and a potentially ugly working copy that allows the use of landmarks, building names etc. for geocoding and analyses.

The third trick is to exploit whatever information source is available to ensure the quality of the work, but all those concerned with the topic being discussed here already have a commitment to this third trick.

Defining Small Areas

Small areas should by definition refer to areas that are places within the city that have both a length and a width and therefore are areas. The issues and concerns to be discussed are for places that are bigger than a house or an apartment building or a parcel lot. They will focus on city blocks which are defined by the Census as "island blocks," that is, areas of land that are bounded by roads, streams, or physical boundaries on all sides. Figure 1 is a simple map of nine island blocks and their boundaries.

Small Areas, Block Faces, Blocks, Places and Social Interactions

Fortunately, the tricks are also relevant to face blocks which researchers, such as Ralph Taylor (1997), felt were important. While Taylor correctly emphasized the limited physical range of the interpersonal dynamics of life, he, in my view, had been a bit more restrictive than would be useful for studying the dynamics of crime. Many of the social processes that he described also can be applied to the places around the corners and on the backs of the face blocks. While there has been little research on this, the existing work as well as the logic of social interaction has indicated the relevance of "island" rather than "face" blocks. William H. Whyte (1956) in The Organization Man noted that social interaction in a suburban environment tended to occur along the sides of the street and across the backyards of the houses. In this environment, the yards in the backs of the houses were not fenced and the yards in the backs of houses on the opposite sides of the block ran into each other without any intervening passageway such as an alley. With this arrangement, it was easy and frequent for children to run from behind their houses to the backs of houses on the other side of the blocks. Interactions also were more frequent down the sides of the streets rather than across the street, as might be assumed. This

mistaken presumption about networks extending across streets probably arose because the residents of the housing facing each other across the street would, in many cases, frequently be in easy view of each other. This perception is, however, an assumption that is not supported by actual patterns of interaction.

Until children reach almost kindergarten age, they rarely have been allowed to cross streets. Prior to this they have usually been rolled down a side of the street in a stroller and then only later are they allowed to stroll down the side of the street to play with neighbors' children, but not to cross the street. In rolling and strolling, the ususal pattern that can be found has involved people walking down a side of a street to a corner before either going around the block or crossing the street to the other side of the street facing individuals' residences.

Interestingly enough, the same patterns were found in Chicago, by one of the most well-known crime pattern researchers, Dr. Carolyn Rebecca Block, who told me personally that, in her Master of Arts thesis, she found that patterns of interaction in Chicago that were the same as Whyte found in his suburban environment. In Chicago, there were alleys separating the backs of houses on the same island block and the garages were located at the backs of the yards instead of being a major part of the front facade of houses. In this type of environment, individuals came to be able to identify their neighbors across the alleys in the backs of their houses as they either pulled their cars into the garages or put garbage in the garbage cans that were out of sight in the alleys rather than putting garbage cans out on the face of the street once a week as is done in many suburban environments or permanently as in some older, more congested cities. In addition to these broad patterns of social behavior, there were crime-related processes whose operations revolved around the island block and not the face block.

In FRONT, In BACK, On the SIDE, AROUND the CORNER, and, later, ACROSS the STREET

Advantages of Small Areas

The major advantage of using small areas has been precision. Although rarely considered in many studies of urban crime and related phenomena such as the work of Robert Sampson and his colleagues (Morenoff et al., 2001) and Robert Bursik (1999), the age-old problem of aggregation error which referred to the averaging of the values of a dependent variable over the subareas of a larger area has plagued urban crime research. While the initial controversy arose over using data for states to make inferences about individuals, all the same issues can be raised for averaging the number of crimes across the city blocks in a neighborhood. Within any neighborhood, no matter how crime-ridden it has been overall, there were always some places that were better than others and some places that were worse than others.

This imprecision can have led to misleading inferences regarding the effects of the characteristics of neighborhoods, although the primary problem will usually have been that the imprecision will have resulted in stronger effects for the characteristics (independent variables) of these larger areas than when using more precise data. The incorrect inferences about the size and strength of relationships also can have affected implications for policy. For example, paying equal attention to all the areas within a neighborhood usually would not make much sense. Fortunately, the officers on the street pretty much have known this and, through experience, learned to allocate differentially their attentions and efforts, but this has been a learning process which those of us on the research end might be able to shorten somewhat. Figure 2 is a map of how crime varied across the blocks in a census tract.

Around and Around Again

Another consideration which was important in considering island blocks was the mobility of behavior that can be involved in the commission of crimes. For an example, I will draw on one of my pet topics -- the relationships between bars and crime. At times, brawls have been known to break out in bars despite the best efforts of the owners or managers, but the number one operational principle of bar owners or managers who would like to retain their businesses or jobs has been “to take it outside” – get if off of my property. If the bar “manager” was successful, the criminal event which began on his/her premises was no longer there and the linkage between the bar parcel and the assault has been eliminated. If the assailant was not completely stewed nor a complete idiot, the best place to have finished the dispute was usually not on the face block of the street with the bar on it. Even neighborhood bars have tended to be on main streets which also tended to be well lit and will often have had police cars driving down them. The best alternative for finishing a disagreement, in many cases, was to have waited for the intended victim to have gone around the block onto a side-street or into a parking lot, if one was nearby. A dark street will usually have been a much better place to have assaulted someone. Thus, a crime has moved not only from the parcel on which a bar was located, but also from the face-block on which the bar was located and the crime began.

Similar considerations also would have applied to robbing a patron coming from a bar. The best place again would have been off the main street on a side street around the corner, around the block, across from the island block on the face block of the side street opposite the side street which was usually at right angles to the face street of the bar. Finally, consider auto

theft, or even assault, or even robbery. Remember, bars again have tended to be on the main streets when they have not been "malled" as in the western part of Omaha. If a potential patron had to drive to the local "watering hole," the best way to have done so would have been to drive so that the car was on the right side of the street just in case there would have been a parking spot on the face block on which the bar was located. If not, and a spot was not obvious on the next face block, as it often would not be, the next best strategy would have been to turn right and to have hoped for a spot on the side street that formed the corner with the face street of the bar block. From this position, the easiest place to have seen a parking spot would have been on the left side of the block that is catercorner from the side street onto which a driver has turned. This side would have been easiest to see because it would be easiest for a driver to have been looking out of the left side of the windshield and the driver's side window. It would have been much harder to look to the right at the back side of the face block with the bar. Turning around to park on the opposite side of the side street which would be on the next block would normally have been difficult. Given a disagreement in the bar that spilled outside, this catercorner block would be one potential location for extending this disagreement. It would also be a good place to rob the driver and, of course, the site of the vehicle theft if it occurred. In short, face blocks can, in very common instances, have been locations that were too small to be the units of analysis when examining the association of crime with the characteristics of areas within the city.

Even a face-block has not been the smallest place used for the analysis of the locations of crime. The work of John Eck (1998) has attempted to use even more micro units of analysis by examining which units within a building were and will be the sites of crime. There has been much value to this work for helping police when they were considering entering a building and

did not have complete knowledge of which unit to approach, but the mobility discussed above for face blocks would be even a more serious consideration for the parts of a building or parcel. In short, in dealing with small areas, the processes likely to be associated with very common crimes of concern to criminologists have made the island block and its neighboring island blocks important areas to consider and they are the ones to which I will pay attention in what follows.

A danger in analyzing small areas, which will also apply to city blocks, has been that crimes across any of the streets which bounded a particular city block would not be counted in that block. Thus, relying only on small areas such as island blocks can have resulted in missing a "hot spot." The same problem, however, also would apply to face blocks. Crimes committed around a corner, in an alley, or on the back side of a street would also not be counted as occurring on a face block and, thus, even relying on face blocks could result in not identifying a "hot spot." Fortunately, potential "hot spots" can be identified from the locations of crimes by using additional methods. The most well-known methods have been included in CrimeStat II (Ned Levine and Associates 2002) and these should be used in conjunction with the analyses of areal data to avoid this problem.

Another behavior pattern that supports the relevance of island blocks comes from noting the usual movement of patrol cars. For the most part, officers driving squad cars will be moving down a main street and when they want to patrol the interior of a neighborhood, right turns to go around blocks will ordinarily be more common than left turns unless it is necessary to circumnavigate the remainder of their beat areas. Good patrol practice will also involve more than merely driving around the boundaries of a beat. It will also involve driving through the beat and this will mean driving around island blocks not just past face blocks.

A great deal of space and discussion has been devoted to justification, but this was necessary. Without it some of the traps and tricks would not make much sense.

TRAPS AND TRICKS

Bad Addresses, but Good results (maybe)

The first trap when dealing with small areas is concerned with the first variant of the inaccurate recording of addresses. There has, at times, been a good practice which appeared to be a bad strategy, but actually was not. In particular, incidents may have been recorded at a corner by using an address that did not exist. The most common one in my experience has been an address that ended with two zeroes, such as 4400 Somewhere St.

There may not have been any visible address on Somewhere St that was 4400 because the corner of 44th and Somewhere was a vacant lot on which a victim was assaulted. The last address on Somewhere St before crossing 44th St may have been 4384 or 4398 and either would have been fine just as long as no address in forty-four hundred range occurred before crossing 44th St.

The first recorded address in the forty-four hundred block of Somewhere St may have been something like 4410 and, in some places, the first legal lot may have had an address like 4402 and not 4400. If there was no other identifiable information on the description of the place of the incident, then the only real solution would be to change the working copy of the street file so that the low address on the even side of the street began at 4400. This may not have corresponded to reality, but lacking other information, it would be the best that can be done.

The blessing of this procedure is that if this kind of problem occurred once, it can be virtually guaranteed that it will occur again. Once a working copy of the map has been fixed to correct this problem, it will be fixed for all time. This is a much better strategy, in my view, than letting software do the thinking and just assigning the crime to the nearest address or calling the address unmatchable again and again and having to make manual fixes or, worse yet, to be presented with selections from a pick list again and again when the pick list did not reflect this incident being a crime on the corner of a legitimate block.

There are some disadvantages to this approach and there are also strategies that can be used to handle these. The first disadvantage is that, in the recording of a crime at this nonexistent address, the recorder/officer may have not respected the odd/even addresses on Somewhere St and may have listed the crime as occurring at 4400, that is on the even side of the street when it actually occurred on the odd side of the street. The first strategy for coping with this potential problem is to use the location codes that most departments, in one form or another, record to identify the type of location in which the incident occurred.

In one city with whose data I have worked, the number “22" would be coded into the field devoted to location to identify a vacant lot and a “17" to indicate a residence. If there was a reasonable correspondence between the location code and the nonexistent address, then assigning the crime to the changed street address in the street file would generally be acceptable. If the location code indicated a house, store, or other building and the address did not exist, then this would be a discrepancy. The first strategy would be to check for the existence of the type of location on the other side of the street. If such a location was on the other side of the street, then a reasonable strategy would be to change the address of the incident. If this were to be done,

however, a special field in the crime record should be created to indicate with a 1/0 variable that an address was being changed and then another field should be created for indicating which kind of change was made. I have an entire series of codes to indicate which types of changes were made. After all crimes have been geocoded, the sum of the dummy variable for an address change can be tallied and used as an independent variable to identify the effect of these changes and to adjust the effects of the other independent variables for having to have made such changes.

The best strategy, of course, especially if one were living in a different city from the one which was the site of the research, would be to visit physically the anomalous addresses. In days gone by, Dick (Richard L.)Block and I drove to all the anomalous addresses of homicides in Chicago and I did as much as possible to fix the maps. As Eric Jefferis, a former grant monitor at NIJ, has discovered, I will still do on-site visits even in Cleveland, if given the opportunity. I have sorely missed doing site inspections in San Diego, especially the western parts. It was always hard to believe how many anomalous addresses turned out to be on beaches.

From experience, problems that were detected for one batch of crime data have usually surfaced again and again, so onsite inspection has been a very efficient strategy for fixing such problems. Another alternative, which has sometimes been useful, has involved checking the assessor's file or parcel file. This has not always been effective because the address assigned to a parcel can depend on which way the parcel was facing. The incident may have occurred on the side of the parcel for which there was no real address, and thus, the parcel file did not always have the real answer. Regardless, it has been a file that I have been most grateful to have received for the first time. Such files have been a blessing and would have helped me so much in

the past if I had them in a form that was as convenient as the compact discs that have become available today.

Fixing a map so that places without real addresses were given segments on a map will save a great deal of work later. In my work with Cleveland maps, I inserted line segments into the working centerline file for major buildings at which crimes were frequently recorded, such as, The Center Building, Terminal Tower, Bond Court, The Arcade, the Galleria. Sometimes these maps will be attached to an existing node or street end, but I have also used segments which essentially floated in the middle of blocks. This may not be possible in software other than the one that I have used most frequently. For the multistory buildings, I assigned address ranges which were consistent with the floors and rooms in a building so that, if a room number was identified in a crime report, it could be matched to these segments. The same strategy can be used for parks, beaches, or other landmarks at which crimes will be recorded but for which "official" addresses did not exist. Figure 3 has an example of a segment attached to a centerline street file to identify The Center Building which was a landmark at which numerous crimes were recorded.

A potential disadvantage of map alteration to cope with poorly recorded addresses was that this strategy could have impaired the linking of such files as aerial photographs to a street map. The lowest address on a block could have been 4410 and this could have been a corner building. Altering the map to have allowed an address of 4400 would have pushed the building closer to the middle of the block on which it stood. This has not been a serious problem when the second principle above was followed which was to keep multiple copies of the maps. Among the purposes for keeping multiple copies of the electronic street maps was that different

maps have to be used for different purposes. Those used for display or for linking with other files should not have contained unusual segments or anomalous addresses. Those used for geocoding work need to have as many as possible anomalous addresses, vanity names, and other oddball segments included in them.

One trap whose bite, at first, might have appeared to hurt, but was not very serious concerned the use of the tally of dummy variables for crime location alteration, i.e., changing the addresses of crimes. Typically, such variables when tallied for small areas such as block groups and then used as independent variables in analyses have been likely to have had statistically significant and possibly strong effects. This can become somewhat disconcerting if key independent variables derived from criminological theory did not have effects as strong as those from the address correction tallies. On the other hand, strong effects from a correction tally variable will have prevented overestimating the effects of key independent variables and having recommended interventions based on these overestimated effects which can have affected the sometimes fragile acceptance of research findings. Also, the finding of strong effects for a correction tally variable can be used to make a strong argument for providing a basis for encouraging and even initiating training in address-recording procedures. The best experience in fostering the encouragement of good recording that I have been told came from Richard L. Block. I recall that he told me that officers became much more accurate about recording the locations of incidents after they started receiving maps of incidents at roll call. Officers whose address recording was not adequate did not have the incidents they recorded on the maps. When they complained, they were told that their address recording was not adequate. As I recall, Dr. Block telling me, these officers then began dramatically improving their recording practices.

BALDing Crimes at Intersections

The second and very major nightmare of working with small areas arises from the tendencies of officers to record crimes at intersections. Very few crimes except a handful of vehicular homicides occur at intersections. When working with large (crude) areas as census tracts or census block groups which could contain anywhere from 1 to 99 census island blocks, intersections could be a serious annoyance, but they will be not nearly as painful as when "playing with blocks" – (I wish my kids wouldn't tell their classes that is what I do for a living). At the census tract or census block group level, this problem of inaccurate reporting only becomes serious when it occurs at a tract or block group boundary. Most of the intersections for these large (crude) units of analysis are internal to these areas, so the majority of crimes even when recorded at intersections are probably put in places which are correct for these larger areas.

This is virtually never true for census blocks. Even when a cul-de-sac is involved, an intersection is still likely to be on a block boundary. Unfortunately, in departments where mapping is a new activity or where there has been little training and encouragement for careful recording and little reward for doing so and no or few penalties for not recording accurately, the number and percentage of incidents that officers record at intersections can easily reach into the triple digits for the frequencies and double digits for the percentages of crimes recorded for the city as a whole. At these high volumes and percentages, simply discarding these incidents seems to me to be a very poor decision, although this might be something that one would do on roll call maps using a version of an electronic map that has not been corrected for anomalous recording just to encourage more accuracy.

For planning and most importantly for research, leaving out ten percent or more of the crimes or even seven eight or nine percent of the crimes just because they are recorded at intersections is not advisable because it can seriously distort the picture of crime. From a statistical standpoint, this is nonrandom error in the dependent variable and will result in biased coefficients whether one is using regression, Tobit, logit, multi-level modeling or any such technique. Also, just simply discarding crimes recorded at intersections is throwing away information. There is partial address information that could be used. In formal terms, the crime recorded at an intersection is prior information. So there is more knowledge about the crime than found in some records in which the address is listed as SOMEWHERE IN THE CITY.

For crimes recorded at intersections, the parts of the city in which the crimes occurred are actually fairly narrowly specified, although they are not specified enough to decide on which blocks in the city these crimes belong and this is an important limitation. The strategy I recommend for dealing with crimes recorded at intersections when the research task is to conduct a block-level analysis is the one to which I and my present students and a former graduate student have assigned the acronym BALD.

BALD is short for Bayesian Address Location Determination. The procedure is Bayesian because it uses prior information and it uses this information to determine the location of addresses, hence BALD. One of the technical requirements of Ordinary Least Squares and other related general linear model techniques is that the dependent and independent variables are measured without error. This is an impossible task in any social science discipline. The consequences of violating this assumption, however, depend on how it is violated. As Lewis-Beck (1980) notes in his Sage Pamphlet on Applied Regression, random error in the

measurement of the dependent variable does not bias the effects of the coefficients in Ordinary Least Squares and this generalizes to other techniques as well.

The core of the BALD technique is to create consciously random error in the dependent variable for which the consequences are known and can be measured rather than producing an unknown bias in the effects of the independent variables. The consequences of such error in the measurement of the dependent variable are a loss of efficiency and the degradation of the minimum variance property of these techniques. In more common terms, the first consequence is that the value of any measure of explained variance for any analysis technique that is used, whether it is a true or adjusted R-squared or a Pseudo-R-squared, will be smaller than it would be without such random error. The second problem that emerges is that the standard errors of the coefficients will be larger than they would be without any measurement error. Thus, it will be harder to obtain statistical significance for the effects of the independent variables. On the other hand, one of the most important aspects of research about which to be careful is not to overestimate the effects of any independent variables in this case, block characteristics. Underestimating is a less serious problem. If some policy based on research produces better results than expected this usually does not impugn the reputation of a researcher.

Although there might be more complicated and computer-based ways of doing BALD, a simple procedure seems adequate. For a crime that is recorded at an intersection, the number of possible blocks on which it could have occurred is usually between two and six. The most common number of blocks in my experience is four. The key procedure in BALD is randomly assigning the intersection to one of the blocks that form the intersection. Any crime incident that is randomly assigned to a block is given a score of 1 on a special field or variable that indicates

the incident has been randomly assigned. When the crimes are aggregated to blocks, this field is also aggregated so that each block has a variable or characteristic that identifies the number of crimes of each type that has been randomly assigned to each block. This variable then becomes an independent variable in any statistical analysis.

The unstandardized effect of these tallies of crime addresses that are randomly assigned to blocks directly measures how much of the variation in a crime is due to the strategy of random assignment and the standardized effect of these tallies measures how important this random assignment was relative to the actual characteristics of the blocks. Including these tallies of random assignments as an independent variable also controls the effects of the other characteristics of the areas for the potential error introduced by having inaccurately recorded addresses that had to be randomized to a location. The size of this effect and its importance are also useful pieces of information to bring to the attention of officials in charge of reporting and data quality to emphasize how much the very poor and inaccurate practice of recording crimes at intersections affects the analysis of where crimes occur. The inaccuracy of recording crimes at intersections could potentially affect litigation against a business owner whose legal advisor challenged the relationship between the crimes and a business because the addresses of the incidents being alleged to involve the business were not recorded at the address of the business.

Before delving into a final technical issue, it will be useful to explain my recommended procedure for conducting the randomization. Unless shown evidence to the contrary, a manual process seems adequate since the results of the process will be identified in the statistical analysis. The first step is to number the blocks at the intersection. There is no apparent reason for using any particular strategy or changing strategies for each crime that is recorded at an

intersection. My strategy is to number the blocks clockwise beginning from the northernmost or northwesternmost block. The next step is to use an MRD (Manual Randomization Device) which can be thrown against a backstop to indicate randomly a number between 1 and 6. The most common such device is a cube that has six sides with each side having from one to six dots on it. Should an intersection have more than six blocks meeting at it, two MRDs or special multisided MRDs as used in role-playing games can be used. There is no apparent reason that using dice to randomize would adversely affect any statistical outcome in a deleterious way.

The final key issue in implementing BALD concerns which blocks at intersections should be considered for randomization. My initial inclinations on this issue were to randomize to only the residential blocks to help preserve the linking of these intersection incidents with demographic, housing, and land use variables. The argument of my former graduate student and now Professor Marc L. Swatt, however, convinced me otherwise. His argument was very simple, random means random, lacking any additional knowledge that would have eliminated a block from consideration, all blocks at the intersection have to be considered as possibly being the actual site of a crime. The possibility that some robberies, assaults (aggravated or sexual), personal thefts, auto thefts, and possibly, but hopefully not, homicides become assigned to nonresidential blocks and then these are omitted from the analyses of crime across residential blocks could be unfortunate, but still would not have biased the results of any statistical analyses.

Randomizing intersection crimes to only residential blocks would produce biases in the results of analyses, at the very least, because such blocks arbitrarily are given a higher probability than nonresidential blocks of being crime sites without any evidence to support their selection. Following such a process would make the levels of crime on residential blocks appear

to be higher than otherwise and this could lead to finding stronger effects for the residential characteristics of the blocks than actually occur and again lead to incorrect inferences and possibly incorrect policies.

When there is prior knowledge about the crime, such as for residential burglaries which could not occur on completely nonresidential blocks, then there is adequate justification for limiting the randomization to the residential blocks. It is at this point where the location codes that police in many departments record for the crimes as well as the parcel file can be especially helpful. Otherwise, lacking strong justification for excluding any block at an intersection, including all blocks, residential and nonresidential, in the randomization process for crimes recorded at an intersection is the best strategy when analyzing crime for small areas. Doing so will not bias results and the effects of randomization can be measured and reflected in the analyses.[1,2]

When Rates are not Great

One of the major traps in working with small areas has been the concern over which strategy should be used to adjust for the partial dependence of the amount of crime on the size of the "small areas." While city blocks have been smallest areas available from Census data, they have varied greatly in size. Much public housing was built on what have been called "superblocks." When such public housing was built as in Chicago and also in Cleveland, parts of old neighborhoods which were basically built on a grid plan had the internal network of streets eliminated and the structure of a superblock closely resembled a very large cul-de-sac with sometimes only one entrance for a very large area. Even when there was more than one entrance

the amount of land on the resulting new blocks was far larger than any of the old traditional blocks. Depending on the style of the public housing, townhouse, three-story, or high-rise, the number of people living on one of these blocks became and has remained far larger than the number of people living on ordinary rectangular residential blocks.

In more recent years in Omaha, apartment complexes and residential areas for the very affluent have been built on superblocks that are very large. Indeed, in Omaha as in Las Vegas as well as other cities, there are walled developments that contain such areas. Very large blocks result from the design of these walled developments. The front of one of these large blocks is usually a single street that runs for a long distance because of the very large lots on which expensive houses sit. Because these lots usually are also quite deep, there is a large distance between the front of the block defined by the street and the wall which defines the back of the block and, hence, such blocks are physically very large. In apartment areas, the long streets will again form the faces of the blocks and there will be substantial distances between the backs of the apartment buildings and the wall which forms the boundaries of these areas. Still, the overwhelming number of residential blocks especially in older cities, but also in some of the newer cities, such as San Diego, are relatively small. Quite frequently, the island blocks are between five and seven acres. Nevertheless, the amounts of variation in the physical sizes and the numbers of residents across blocks require adjusting statistical analyses of the frequencies of crime in small areas for these variations.

From Shaw and McKay (1942) onward, a common practice among criminologists has been to follow the convention of adjusting for the varying size of the areas being analyzed statistically by using rates and most unfortunately using population-based rates. When analyzing

the frequencies of where crime incidents occur, particularly for crimes against persons and moveable objects, population-based rates are meaningless even for purposes of testing theory. Individuals and their possessions do not have to be permanent residents of the neighborhoods, census tracts, block groups, city blocks, or face blocks on which they are victimized. Therefore, the number of residents for any one of these areas does not represent the population-at-risk of being victimized. People can be mugged on blocks, in block groups, and in census tracts in which they do not live. Cars can be stolen from areas in which their owners do not live. Worse yet, population-based rates can seriously distort the apparent dangerousness of areas. Just a handful of crimes on a block with a few people living on it can generate a very high crime rate even when these few crimes are committed against the few residents. On the other hand, a large number of crimes in one of the huge apartment complexes, such as a large public housing project, can result in very low rates even though such places might have many times the number of crimes found in the areas with few residents.

Interestingly enough, Osgood (2000) now even raises the issue of whether population-based rates should be computed for large areas, such as cities, because crime is inherently a count variable, i.e., frequency. His strategy which I will not review here, however, still may be inadequate for small areas because it relies on a fixed effect for population. My recommended strategy is to control for the physical size of small areas and for the number of residents explicitly by using the measures of the physical size and number of residents of small areas as independent variables in any analysis. For future urban planning, the results of analyses that explicitly include measures of size as independent variables can inform those making decisions about the shape and size of residential areas about the potential consequences of their decisions.

This is a practice that I have followed from the early 1980s and beyond. Perhaps surprising to some, neither measure of size will completely dominate any analyses for small areas. Other characteristics will still continue to have statistically significant and important effects.

A final technical problem with a population-based rate is that the numerator of the rate, the number of crimes, typically has a much smaller range than the number of residents in small areas and, thus, the number of crimes has a smaller variance than the number of residents. As a result, the numerical values and variance of any population-based rate will be largely determined the number of residents in the areas for which the rate is computed. Therefore, statistical analyses of population-based rates will largely be analyses of the variation in the number of residents living in the areas.

Two other considerations provide credence for treating dependent variables that are inherently frequencies as frequencies and explicitly measuring the effects of the physical and population sizes of small areas. First, there is justification for the theoretical relevance of the number of people in an area from very early urban sociological theory (Wirth 1938) which described how the number of people in a city could produce deleterious consequences. This position has received additional empirical support from much of my past work and has also received recent theoretical support from more recent statements of Routine Activity Theory (Felson 1998). Second, if research on crime in small areas is to have an impact (a noun not a verb) on policy, then this research must address the frequency of victimizations, i.e., how many crimes occur where. Police cannot and crime researchers should not focus on rates when incidents are the problem, focusing on incidents is what is needed for crime prevention and enforcement.

There are three cases when rates, in my view, would have utility. The first two really are just variants of each other. Computing rates for residential burglary based on the number of residences in area could be useful. The number of residences represents what demographers commonly refer to as the "population at risk." The fatal flaw in population-based rates is that the number of people in an area at any point in time is not restricted. Varying numbers of individuals who do not live in an area can be there at almost any time, so that there is no way of knowing how many people were at risk of being a victim of crime at any particular point in time. Unlike people, residences, whether in houses or apartments, tend to be permanent fixtures in an area whose numbers tend not to fluctuate much on a daily or hourly basis. Only rarely do residences come from one area to visit another for only brief periods of time. Residential burglary rates based on the number of residences, thus are one type of valid rate. Commercial burglary rates are the second type of valid rate that could be constructed, but issues are more complex here, despite this being merely a different type of burglary rate. Whether the number of commercial establishments or the square footage devoted to commercial activities would yield more useful rates is not always clear.

The third and final type of valid rate for the risk of crime is one that could potentially be used for crimes against persons in small areas. Rates based on the number of personal crimes committed per acre of land territory can be a legitimate measure of the dangerousness of areas. At least potentially, almost every inch of territory can be at risk of being the location of crime. Individuals can be robbed or assaulted in doorways, pushed through windows, and, particularly with the change from walls made of plaster to those now made of plasterboard, individuals can now be slammed into walls in the process of an assault.

All three rates are valid measures of risk because the denominator of each rate is a fixed and known quantity. Such rates are probably most useful for identifying the types of places which, in principle, are dangerous. The evaluation of risk is important for formulating, testing and developing theories about which characteristics of areas produce crime. Since these rates are true measures of risk, they can be used for this purpose. These rates, however, have two major limitations. First, the variation in the number of residences, number of businesses, and physical sizes of areas will generally exceed the variation in the number of crimes. Thus, it is still the case that analyses explaining these rates are partially explaining the variation in the denominators of the rates. Regardless, because these denominators are the locales that are exposed to the risk of crime, this problem is not serious when compared to the same problem for population-based rates. Second, even these rates are of limited use for short-term policy for law enforcement. A block which has one commercial burglary in a year and two businesses on it has 50% burglary rate. A block with five burglaries but fifteen businesses on it, has only a 33% rate. The police still have to pay more attention to the low-rate block than the high-rate block despite the difference in the values of a valid risk-based rate.

The best strategy, in my view, is to analyze the frequency of crimes and to control for the sizes of the residential populations, the physical sizes of the small areas, as well as their proximity to other areas. The generalized potential is a useful measure for controlling for the overall proximity of any area to all other areas in the surroundings (Roncek and Montgomery 1984; 1995). Such analyses are needed from a practical standpoint and do have a theoretical basis supporting their scientific use. Examining area-based rates for personal crimes and, perhaps, vehicle theft can be useful for scientific purposes in helping to identify how the effects

of the characteristics of areas differ from those which emerge when using the frequencies of crimes as the dependent variables. The usefulness of the results of analyzing risk-based rates for short-term planning is likely to be more limited than the usefulness of the results of the analyses of the frequencies of crimes.

The types of statistical analyses to be used for small areas will be discussed later. Issues that have not been discussed sufficiently because the use of small areas is still relatively rare need to be addressed first. Articles published in the most prestigious journals are still combining census tracts into larger units rather than examining areas that constitute the basic building blocks of neighborhoods.

Data for Here not There

Apart from the extra effort required, one of the rationales for not using small area data in the past has been, in particular, the absence of data on income for city blocks. Given that up to the 2000 Census, measures of the value of owner-occupied housing and measures of the value of rent were available, this rationale has seemed to be a poor one. There were a variety of ways to cope with this problem which were used only rarely. The excuse for using block groups, which were not defined by any meaningful social, housing, or geographic criteria, has been that these were the smallest units for which income data were available. Again, this seemed to be a poor rationale, in my view, given that parcel data have become much more widely available than ever before. In years past, such data were only accessible in the dreams of urban crime researchers. The assessed value of owned housing from assessor or parcel files was much more likely to be a better indicator of the affluence of small areas than any housing value or income data collected

by the Census. Assessed values have been assigned, at least in a seemingly large percentage of cases, by a disinterested party using data rather than by the respondents whose lack of a full understanding of the security of Census data might have led to the misrepresenting of their incomes. Any complaints about the lack of measures of poverty for small areas were essentially the same complaints about the lack of income in different clothes. For education, there never were adequate substitutes at the block-level, but research with larger units of analysis has not shown that any measure of it really mattered for crime once measures of affluence, household structure, facilities, public housing, and age compositions were taken into account.

The lack of data for city blocks in the 2000 Census is more serious than ever before. Measures of the value of owned housing, the value of rent, the types of apartment buildings, and, most surprisingly, for an aging population, the measures of the number of condominiums are no longer reported. Fortunately, the availability of parcel data can compensate for much of the valuable information that is no longer in the 100% Census questionnaire for 2000. The most difficult characteristic to estimate with 2000 data is a measure of affluence for blocks or areas that have only rental units. At this point, no easy and satisfying solution is readily available. The easiest solution would be to substitute a measure from the block group data such as the average assessed value of owned housing and then to set a dummy variable indicator to identify for which blocks this substitution was made. If necessary, the tract-level value could be used. Another solution would be to estimate an equation for housing value based on the other characteristics of the blocks including their relative locations in the cities and then to use this predicted value for the blocks without a value of owned housing in the parcel file. There might be a small risk of multicollinearity, but this seems unlikely since this estimation would probably

only be necessary for a small percentage of the blocks.

A solution that would seem apparent to many would be to use one of the Hierarchical Linear Modeling programs to estimate the effects of characteristics that are missing from block-level data that are available at a higher level of aggregation. Potentially, this would allow including measures of income and education in the analyses of crime in small areas. This possibility has not been discussed to my knowledge and this strategy could still have some limitations. In using multi-level modeling, investigators must identify how the characteristics of the larger units of analysis affect the lower-level units. The higher-level characteristics can either affect the coefficients of characteristics at the lower level or affect the intercept of the lower-level equation.

Two issues arise in attempting to use this strategy. First, there does not seem to be any developed theory which would guide researchers in arguing for a tract-level or block group-level characteristic affecting the coefficient of a block-level characteristic. For example, it is not clear how to argue that the average income of the census tract or block group would affect the coefficient of the number of female-headed families with children under the age of 18 or the percentage of persons of age 65 unless the issue of time-order could be addressed. Causation would seem to run in the other direction with, generally, the number of female-headed families on the blocks affecting the average level of census tract income.

Another alternative would be to allow the characteristics at the higher-level of aggregation affect the intercept of the block-level or smaller area equation. It might make sense to argue that blocks located in different census tracts have different initial levels of crime depending on the broader environment. Using this strategy requires arguing that characteristics

at the broader level set the baseline for those at the lower level. Such an argument might be made for the average income at the census tract level affecting the intercept in the block-level crime equation, but this strategy seems to require more supportive arguments than can be developed in this discussion. This might be a reasonable strategy, but it is not clear whether the higher-level characteristics apply equally well to both interior blocks and blocks on the borders of the census tract. Potential arguments for using block group or census tract characteristics are that block groups are subareas of census tracts and that census tracts are defined using homogeneity criteria. Also, for blocks on the borders of census tracts, creating pseudo-tracts by aggregating the block data for bordering blocks would not permit creating substitute measures for characteristics that are not already in the block data.

At the moment, I am working on an additional solution, which I call contextual residualization. This technique involves using the block-level data to predict the value of block group and census tract characteristics that are not reported for blocks. The residual for each block is the amount of the block group or census tract characteristic that cannot be accounted for by the characteristics of a particular block. Because this measure of each higher-level characteristic is a residual from the block characteristics, it is completely uncorrelated with the block-level characteristics that were included in the equation to predict the higher-level variable. This residual may then be included in a block-level analysis equation for explaining crime. These residuals then measure the effects of block group or tract-level characteristics on block-level crime and these residuals reflect influences that cannot be accounted by the characteristics of each of the blocks.

In essence, this strategy is somewhat Durkheimian in that it is using characteristics of the

whole (block groups or census tracts) that are something more than the parts themselves. I have yet to compare the results of this strategy with the results of multi-level modeling, but my intuitive sense is that this procedure may be better because it does not require the specification of effects on either the coefficients of the level-1 independent variables or on the level-1 intercept.

Crime in the Surroundings

For small areas, the effects of crime in the surroundings could be more important than for larger areas of analysis, if for no other reason, then as described in the old Geography adage, things which are close to each other affect each other and the closer they are the more they affect each other. Thus, it may be more important to identify spatial autocorrelation effects for small areas than for larger areas.

Working with small areas can be much more difficult than working with larger areas because as the size of the areas decrease, e.g., down to city blocks, there will be other small areas which do not belong in the analyses (nonresidential blocks) from which crime can spread to residential blocks. When census tracts or aggregates of census tracts or larger units such as counties are the units of analysis for examining differences in crime levels or rates, it is quite rare to find any such areas that do not have residents and, therefore, do not have values for the demographic and housing variables to be used in the analyses.

When using smaller units of analyses, perhaps even at the level of the block group, there can be totally nonresidential areas such as shopping malls, industrial areas, commercial areas that have no residents and no housing but can still be the sites of crime. Simply including these places in analyses and letting these places have values of zero for the percentage of African-

Americans or for housing value could have very deleterious results on the analyses. A place in which no African-Americans are living is not necessarily in an all-white, or all-Asian area, it could be surrounded by an entirely African-American community. A similar serious distortion would occur from allowing housing value to have a value of zero for blocks without any residents or without any owned housing. Such a strategy would seriously distort the mean, variance, and covariances of housing value and, thereby, produce misleading results.

Crimes on these nonresidential blocks could conceivably spill over and affect the amounts of crime on the bordering residential blocks. Conversely, if these nonresidential places are not the sites of crime, then it is possible that being next to such areas might result in less crime on the residential blocks next to these blocks than on other residential blocks which are bordered solely by residential blocks. In any case, the influences of these places need to be taken into account. All existing procedures including SpaceStat (Anselin 1992) and now GeoDa (Anselin 2004) appear only to allow spatial autocorrelation effects from units of analysis that are included in the analyses. With Marc L. Swatt, I am now working on a strategy based on the generalized potential that we believe will allow these effects to be incorporated. We will be presenting the details of our formulation at this year's American Society of Criminology meetings.

Two other problems arise in examining spatial autocorrelation effects. First, it appears that in both SpaceStat and GeoDa, it is necessary to identify the range of the spatial autocorrelation effects. Thus, it may be necessary to compute and recompute analyses with different levels of proximity ranging from the areas which immediately border each of the small areas to those which border the bordering blocks on outward until no effects are found. There is

no guarantee that the only crime diffusion effects will come from the immediately surrounding blocks, although this has been the pattern found in my past research.

Second, and this may be a problem with the functioning of the operator, i.e., me, I have not been able to estimate spatial autocorrelation effects with SpaceStat for the number of blocks in the city with whose data I am currently working. SpaceStat simply stops running when I attempt to use it for the 7,000 blocks in my city. Apparently, GeoDa will handle this many and many more cases, but I have not had a chance to use the latest version which computes a spatial regression. At the moment, it would appear that GeoDa may be the computer program, of choice, although as Caterina Gouvis-Roman's dissertation (2002) notes, there are still not readily available computer programs that simultaneously adjust for spatial autocorrelation and that parallel the Poisson-family of regressions in terms of being appropriate for count data. Relying on GeoDa, may be the best solution for the moment. My graduate students and I are experimenting with techniques used in another field at this point as well as a solution based on my generalized potential (Roncek and Montgomery 1984; 1995), but we are not yet ready to advocate fully their use as a better alternative.

Going Around in Circles: Radius vs. Adjacency

A common concern of a substantial amount of research is with crime within the vicinity of facilities. Part of this concern may have been generated by legislation creating drug-free zones around schools. The measurement strategy is typically to use a circle of a specified radius around a particular facility, such as schools and most recently bars. While legislators defined drug-free zones using radii, probably because they were simple, circular areas do not reflect the

topography of the city and can distort findings with regard to the diffusion of crime from facilities. This type of research seems to be motivated more by what the software can do rather than by what should be done.

Apart from some of the needed efforts to evaluate the effects of policies based on radii, using circles can miss important aspects of criminal behavior that could be related to a facility. Recently, I and one of my graduate students (Roncek and Murray 2003) compared the importance of the diffusion effects of bars on only those assaults within a 500-foot radius of a bar block and on the total of all the assaults on blocks next to blocks with bars. Using only the radius omitted substantial numbers of crimes from the analyses and produced a negative effect on the crimes within this radius. In contrast, being a city block adjacent to a bar block had a statistically significant and important effect on all the assaults on the adjacent blocks.

In short, using circles cuts off important areas of nearby blocks where crimes did occur which were statistically related to the proximity to a bar. Crimes of several types including drug deals do occur at corners. Circular buffers can omit the corners of blocks and the crimes on them from the tallies of crimes in these buffers. In short, it appears that misleading inferences occur from letting the software do the thinking. This circular shortcut which cuts off corners and cuts blocks short is really not necessary. The query procedures now in GIS packages permit identifying which areas are next to which others and permit creating variables that reflect the characteristics of the entire blocks which border those of most concern for analysis from these identification procedures. The recommendation for examining the effects of proximity to a facility is not to be trapped by the apparent ease of a circular buffer procedure, and the trick is to use the query capability to respect the topography of the city when examining diffusion effects.

Statistical Analyses for Small Areas

Inherently, crime data are count data and, generally speaking, the proper multivariate analysis strategies for count data will come from the Poisson regression family. Currently, if an adjustment is to be made for spatial autocorrelation, using one of these strategies is not possible. The current multivariate techniques which are available and are capable of adjusting for spatial autocorrelation bear more similarity to Ordinary Least Squares procedures than to Poisson-type regressions. Should spatial autocorrelation effects prove not to be statistically significant, then Poisson-type analyses can be used, but it may be important to raise the issue of when it is best to use them.

As the size of the areal unit of analysis gets smaller, the range of values taken by any type of crime gets smaller. At the extreme, usually murder, there may be no more than three or four murders on a block, if that (Roncek et al. 1988; Roncek et al. 1991). At this point, Poisson regression analyses are often not useful. An ordinal logit is likely to be the best procedure to use. Indeed, for homicide, using a dichotomous dependent variable, murder–no murder, might be the best that could be done because there can be so few areas with more than one murder. Should the frequency distribution of a particular type of crime be somewhat denser but still quite restricted, and there are no strict guidelines on this, although a highest value of five or six with perhaps at least a dozen cases with the highest values would usually be a good rule of thumb, then ordinal logit would generally be the best technique to use. Poisson/negative binomial regressions, and Tobit usually do not converge to a solution for the coefficients as quickly with this small range or may not converge at all. Using ordinary regression on a dependent variable

with such a small range and many cases with values of zero is very likely to produce predicted values which are far out of the range of the actual amounts of crime across the blocks or other small areas.

For an intermediate range of values for which the highest value is generally not more than once a week, i.e., 52 crimes a year, the Poisson family of regressions seems most appropriate. Crimes are inherently count data and these are the data for which Poisson procedures were designed. With an extremely large range of values that substantially exceed 52 crimes a year, the Poisson family of regressions tends to underestimate substantially the amount of crime in the highest crime places (Roncek 2000). This error is extremely dangerous when attempting to make results relevant to policy. When the maximum frequency of crime begins to approach or exceed one crime per week, the distribution of crime begins to approach continuity. At these levels, spatial autocorrelation regressions begin to become more appropriate. The major problem which will remain concerns the percentage of cases that have zero as a value. If this is small, then the standard spatial regression approach is viable. If the percentage is large, then a censored dependent variable technique, such as Tobit, could be more appropriate. Unfortunately, logit and Tobit computer programs which adjust for spatial autocorrelation are not readily available at this time. Currently, my work with my now new colleague on an alternative procedure for measuring spatial autocorrelation effects should, in principle, allow making the necessary adjustments to any general linear model estimation strategy.

One final issue deserves mention. The residuals from an Ordinary Least Squares regression are simply the difference between the predicted values and the observed values of the dependent variable and there is a complete consensus on this calculation. Such a consensus does

not exist for techniques based on maximum likelihood estimation. There are a variety of residuals which can be computed including Pearson residuals and deviance residuals. In terms of usefulness for policy, the basic question a practitioner is entitled to ask to a researcher is "How close were your predicted amounts of crime from your statistics to the actual amounts that occurred?" Answering this question simply and forthrightly requires nothing more than identifying how close the predicted values were to the actual ones and identifying the error as simply the difference between the best estimate for each area and the actual amount of crime in each area. This may be statistical heresy, in some sense, but the data themselves and criminals and victims do not care whether the residuals should be weighted to account for overdispersion at high values of the dependent variable. The concern of potential victims is more likely to be with how accurate researchers have been in estimating the accuracy of the amount of crime in different areas.

SUMMARY and DISCUSSION

The problems inherent in working with small area data are more difficult than those encountered when working with large units of analysis, such as block groups, census tracts, or aggregates of census tracts. These problems, however, are not so severe that the gain in precision and policy relevance from using small areas should be surrendered for ease of computation and data manipulation. There are traps when working with small areas. The major ones discussed were artificial addresses, the chronic problem of intersections, the lack of data for variables which are (often without evidence) regarded as important, the trap of population-based rates – i.e., non-risk-based rates and rates in general, the increased importance of diffusion and

spatial autocorrelation effects, the danger of going around in circles and ignoring the topography of a city as well as the locations of crimes, and the rareness of the frequency distributions of crimes for small areas which call for the use of techniques that have not yet been incorporated into spatial analysis statistical models. Fortunately, for all of these traps, there are tricks which can provide solutions or, at least "workarounds" for these potential problems. These tricks are easy enough to use, for the most part, so that little excuse should remain for continuing to wander aimlessly or in circles through approximations to neighborhoods or areas designed for the convenience of census enumerators expecting never to be safer and never to be more in danger than the average amount of crime for a larger area indicates regardless of the part of a neighborhood in which the wandering is occurring.

In closing, after this lengthy discussion, all the advice and argument can be pulled together in what could "humbly" be called Roncek's Recommendations.

1. Use island blocks, not block groups, nor census tracts and by no means anything bigger unless it is solely for policy purposes.

2. Keep multiple copies of all electronic maps. At least one copy of the centerline street map should be a working copy that has segments for vanity street names, buildings, landmarks, parks, etc. at which crimes occur frequently. Having a working copy with such adjustments still turns out to be important even when crime addresses are processed through a large "scrubber" computer program which repairs known problems in recording before geocoding. Keep one very clean copy for presentation purposes.

3. Use whatever software is appropriate for a particular task. For geocoding, it is most helpful and important to use software which provides codes for why a crime address did not geocode correctly. Always require a 100% exact match on the first attempt to geocode so that the amount of problems in the data can be identified and provided as feedback to those who have the authority to improve the quality of crime address recording.

4. Whenever possible visit the sites of crimes which did not geocode at the 100% criterion to determine the cause of the failure to geocode. Once a problem with a crime address occurs, it is very likely to occur in additional data. A site visit will permit determining the nature of the

problem which can guide the changes made to the electronic map so the problem does not recur and will permit identifying and confirming the characteristics of a particular address so that information from a crime report such as, a location code, can be used to avoid other problems.

5. Allow for crime recorders/police to record addresses of incidents at nonexistent addresses at vacant lots or parcel sides as long as the recorders locate the incident on the correct odd-even side of the street. Check the correctness of the odd-even assignment whenever possible using codes which identify the type of location at which a crime occurred. If necessary, reassign the crime to an address for which the lot or facility at an address matches the location code. Create a field in each crime record and in it identify the type of change that was made to the address of a crime. Create a second field in the crime record that has a 1 in it if a change was made to a crime address because of the location code. This field will be tallied and used as an independent variable to control the effects of other independent variables for corrections that had to made to the crime addresses. Its effects in all subsequent analyses measure the effects and importance of having to alter addresses because of inaccurate reporting. These effects need to be reported in any research documents and to the appropriate authorities.

6. Randomize the crimes recorded at intersections to the blocks at that intersection regardless of whether the blocks are residential or nonresidential. Create a field in the record of each crime to indicate if the block to which the crime was assigned was randomized. Aggregate this field to a field in the block file and use this field as an independent variable in all analyses. Report the tally of each type of crime that was randomized and it effects for any analyses in any research report and to the appropriate authorities

7. Do not use population-based rates for the analyses of crime in small areas within a jurisdiction regardless of whether the areas are blocks, block groups, census tracts, or even larger areas.

8. For strategic and tactical planning, use the frequency of crimes as a dependent variable.

9. For testing theories, using risk-based rates can be valuable especially when the results of analyzing these rates are compared with the results of the analyses of crime frequencies.

a. For personal crimes, thefts, and vehicle thefts rates based on the physical size of small areas are valid risk-based rates. They still have limitations for strategic and tactical planning.

b. For residential burglaries, a valid risk-based rate can be created using the number of housing units as the denominator of the rate. Such rates will still not be fully adequate for enforcement and prevention planning.

c. For commercial burglaries, use either the number of businesses or the square footage devoted to commercial uses as the denominator of a crime rate. Again, even these rates cannot be the only measure of the extent of these problems in different areas.

10. Use any and all sources of data that can be obtained for identifying important

characteristics of small areas which are not readily available for census blocks. Use either multi-level analyses or contextual residualization to identify the effects of characteristics available at a larger unit of analysis that are not available for the small areas being analyzed.

11. Adjust all analyses for the population size and physical size of the blocks using these as independent variables.

12. Adjust all analyses for the level of crime in the surroundings of the small areas, city blocks, block faces, etc.

13. Unless doing a policy evaluation, do not use circular buffers to identify the effects of proximity to a facility such as a bar or school. Examine the effects of proximity using the blocks or small areas that are adjacent to a facility. Crimes and drug deals often occur at corners and these are often not included in circular buffers around facilities.

14. For the statistical analyses of the patterns of very rare crimes, use dichotomous (binomial) logistic regression to identify the effects on the probability of having a rare crime occur.

15. For the statistical analyses of infrequent crimes, use ordinal logit to identify the probabilities of being at the different but very few levels of infrequent crimes.

16. For relative frequent crimes that occur approximately less than once a week in an overall study area such as a city, the Poisson family of regressions including overdispersed Poisson and Negative Binomial Regression would be the most appropriate techniques.

17. For very common crimes that occur more than once a week in the overall study area and have a relatively wide range across the small areas, but for which many blocks or small areas do not have any crime of these types, Tobit analysis would be the preferred analysis strategy.

Notes

1 After the workshop, a participant suggested altering a strategy that I recommended. After some thought, it became clear that my original strategy was the one which should be used. The suggestion was to assign each block at an intersection the probability it had of being the site of a crime that was recorded at an intersection rather than randomly assigning the crime to a block through the BALD procedure. There are two problems with this suggestion. First, regardless of whether one assigns the crime to a randomized block or assigns each block at an intersection the probability of being location of a crime, e.g., at a four-block intersection each block would be given the value of .25 for a particular crime, the number of times a block was assigned a probability for an intersection crime must be tallied and the used as an independent variable. The simplest and most direct way to tally would be to use a 1 every time a probability was assigned to a block just as when a crime is BALDed to a block. Attempting to compute a tally using the probability assigned to a block would make it difficult if not impossible to interpret the tally. For example, a block at a four-way intersection that had four crimes assigned to it would receive a score of 1 on the tally as would be the case for a block at a three-way intersection that had three crimes assigned to it. While it would be possible to talk about this measure as the cumulative probability of changes, the total number of changes made would be lost. This total is particularly important for reporting back to officials who have the authority to improve the quality of address-reporting. The actual tally of the number of changes made is also important because it will provide the most interpretable measure of how much poor recording accounts for the variation in a particular crime type and because the effects of other independent variables need to be controlled for the actual number of assignments made and this is lost when using the probability-based strategy.

Another important drawback of the procedure suggested by the participant is that the probabilities assigned to blocks at an intersection when tallied along with the number of crimes correctly geocoded to the block change the character of crime as a dependent variable. Having some blocks with crime tallies that involve fractions of a crime along with blocks for which the number of crimes is an integer based on the actual count of correctly geocoded crimes or a tally of probabilities for which the total was an integer makes the choice of analysis technique to use unclear at best. Because the values of the crime variable are no longer fully counts or integers, regressions from the Poisson family are no longer appropriate. On the other hand, because these values are not truly continuous, ordinary regression (OLS) is not appropriate. The problem of the lower bound of 0 for the number of crimes will remain for most if not all crime types and the lack of true or even approximate continuity which is unlikely to occur would be a violation of the conditions needed to use limited dependent variable techniques such as logit or Tobit.

2. Another participant questioned the use of the tally of crimes that were BALDed because there would be a tendency for larger blocks to have more crimes BALDed to them than would be the case for smaller blocks. While this could occur, one of the major recommendations made for working with block data is to use measures of size as control variables in all analyses. Following this recommendation would eliminate the problem of having more BALDed crimes on larger city blocks

REFERENCES

Anselin, Luc. 1992. SpaceStat Tutorial: A Workbook for Using SpaceStat in the Analysis of Spatial Data: Urbana-Champaign, Illinois: University of Illinois

Anselin, Luc. 2004. GeoDa. Council for Spatially Integrated Social Science. Urbana-Champaign, Illinois: University of Illinois.

Block, Carolyn Rebecca. 1987. Personal Communication.

Bursik, Robert J., Jr. 1999. "The Informal Control of Crime Through Neighborhood Networks." Sociological Focus 32: 85-97.

Eck, John. 1998. "Preventing Crime by Controlling Drug Dealing on Private Rental Property." Security Journal 11:4.

Felson, Marcus. 1987. "Routine Activities and Crime Prevention in the Developing Metropolis." Criminology 25:911-932.

Felson, Marcus. 1994. Crime in Everyday Life. Thousand Oaks, California: Sage.

Gouvis-Roman, Caterina. 2002. Schools as Generators of Crime. Ph.D. dissertation in Sociology, American University: Washington, D.C.

Lewis-Beck, Michael S. 1980. Applied Regression: An Introduction: Thousand Oaks, California: Sage.

Morenoff, Jeffrey, Robert J. Sampson, Stephen Raudenbusch. 2001. "Neighborhood Inequality, Collective Efficacy, and the Spatial Dynamics of Urban Violence." Criminology 39:517-560.

Ned Levine and Associates. 2002. CrimeStatII. Washington, D.C.: National Institute of Justice.

Osgood, D. Wayne. 2000. "Poisson-based Regression Analysis of Aggregate Crime Rates." Journal of Quantitative Criminology. 16: 21-43.

Roncek, Dennis W. 2000. "The Accuracy of Predicted Values for Count Data: A Comparison of Logged OLS, Tobit, Negative Binomial, and Least Absolute Deviation Regressions." Paper presented to the 2000 annual meetings of the Midwest Sociological Society, Chicago, Illinois.

Roncek, Dennis W., Richard L. Block, and Max Ewalt. "The Ecological Persistence of Homicide in Chicago." Paper presented at the 1988 Annual Meetings of the Midwest Sociological Society in Minneapolis.

Roncek, Dennis W., Vicky R. Anderson-Weatherspoon, Richard L. Block, and Carolyn Rebecca Block. 1991. "The Similarities and Differences Between the Ecological Patterns of Black-on-Black and White-on-White Homicide in Chicago." Paper presented at the 1991 annual meetings of the Midwest Sociological Society in Des Moines, Iowa.

Roncek, Dennis W. and Andrew C. Montgomery. 1986. "Spatial Autocorrelation: Diagnoses and Remedies for Large Samples." Paper presented at the 1986 annual meetings of the Midwest Sociological Society in Des Moines, Iowa.

Roncek Dennis W. and Andrew Montgomery. 1995. Spatial Autocorrelation Revisited: Conceptual Underpinnings and Practical Guidelines for the Use of the Generalized Potential as a Remedy for Spatial Autocorrelation in Large Samples." Pp. 99-110 in C.R. Block and M Dabdoub, and S. Fregly (eds.) Crime Analysis Through Computer Mapping. Washington, D.C.: Police Executive Research Forum.

Roncek, Dennis W. and Rebecca K. Murray. 2003. "A Comparison of Diffusion Effects Using Radius and Adjacency Measurement Strategies." Paper presented at the 2003 annual meetings of the Academy of Criminal Justice Sciences in Boston, Massachusetts.

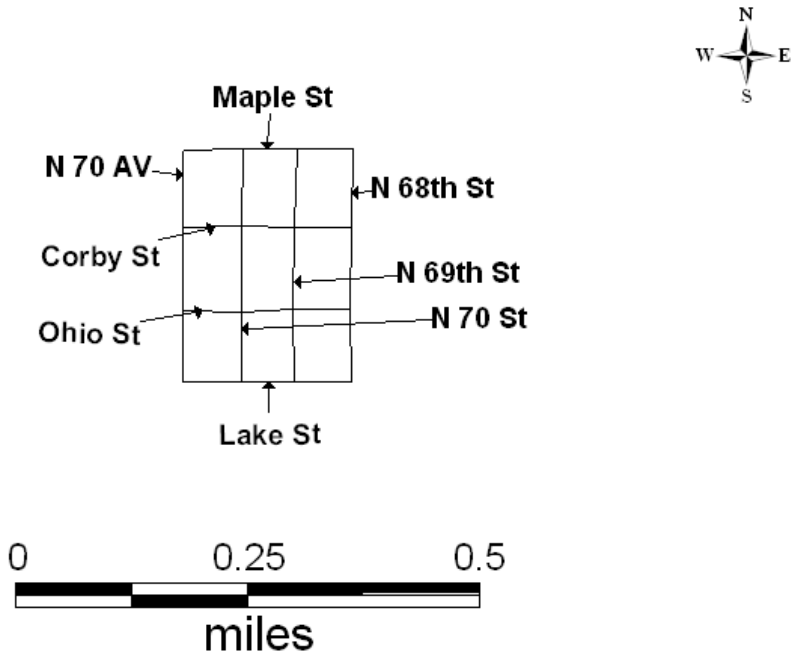
Shaw, Clifford R. and Henry D. McKay 1942. Juvenile Delinquency in Urban Areas. Chicago: University of Chicago Press.

Taylor, Ralph B. 1997. "Social Order and Disorder of Street Blocks and Neighborhoods: Ecology, Microecology and the Systemic Model of Social Disorganization. Journal of Research in Crime and Delinquency 34: 113-155.

Whyte, William H. 1956. Organization Man. New York: Simon and Schuster.

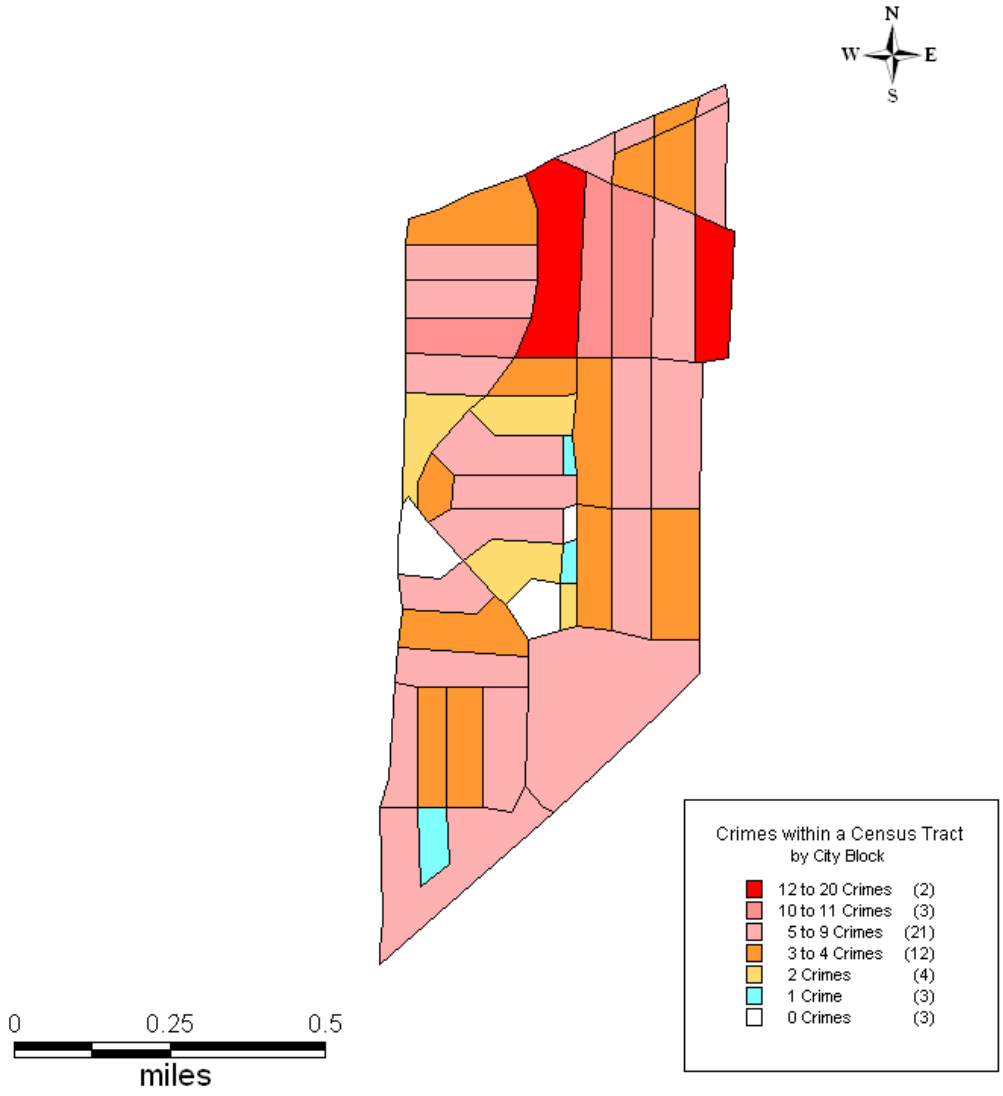
Wirth, Louis 1938. "Urbanism as a Way of Life." American Journal of Sociology. 44: 1-25.

Figure 1. Nine Island Blocks with Street Boundaries



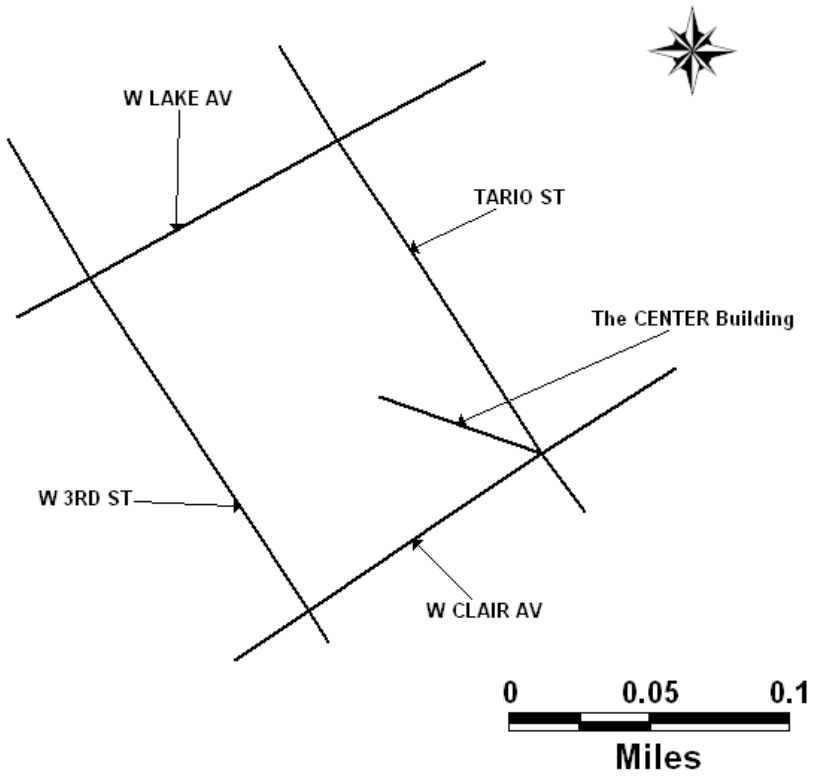
Map by Dr. Dennis W. Roncek, Department of Crimnal Justice, University of Nebraska at Omaha

Figure 2. Variation in Crime Among City Blocks Within a Census Tract



Map by Dr. Dennis W. Roncek, Department of Criminal Justice, University of Nebraska at Omaha

Figure 3. Inserting Landmarks into a Street Map to Enhance Geocoding Success



Map by Dr. Dennis W. Roncek, Department of Criminal Justice, University of Nebraska at Omaha bosbldg04